Proc. of the 5th Int. Conference on Digital Audio Effects (DAFX-02), Hamburg, Germany, September 26-28, 2002

SUB-BAND INDEPENDENT SUBSPACE ANALYSIS FOR DRUM TRANSCRIPTION

Derry FitzGerald, Eugene Coyle

D.I.T., Rathmines Rd, Dublin, Ireland
derryfitzgerald@dit.ie
eugene.coyle@dit.ie

Bob Lawlor

Department of Electronic Engineering,
National University of Ireland, Maynooth
rlawlor@eeng.may.ie

ABSTRACT

While Independent Subspace Analysis provides a means of separating sound sources from a single channel signal, making it an effective tool for drum transcription, it does have a number of problems. Not least of these is that the amount of information required to allow separation of sound sources varies from signal to signal. To overcome this indeterminacy and improve the robustness of transcription an extension of Independent Subspace Analysis to include sub-band processing is proposed. The use of this approach is demonstrated by its application in a simple drum transcription algorithm.

1. INTRODUCTION

1.1. Independent Subspace Analysis

Independent Subspace Analysis (ISA) was first proposed by Casey and Westner as a means of sound source separation from single channel mixtures of sounds [1]. ISA is based on the concept of reducing redundancy in time-frequency representations of signals, and represents sound sources as low dimensional subspaces in the time-frequency plane.

ISA makes a number of assumptions about the nature of the signal and the sound sources present in the signal. The first of these is that the single channel sound mixture signal is assumed to be a sum of p unknown independent sources,

$$s(t) = \sum_{q=1}^{p} s_q(t) \tag{1}$$

Carrying out a Short-Time Fourier Transform (STFT) on the signal and using the magnitudes of the coefficients obtained yields a spectrogram of the signal, $\boldsymbol{Y}$ of dimension $n \times m$, where n is the number of frequency channels, and m is the number of time slices. From this it can be seen that each column of $\boldsymbol{Y}$ contains a vector which represents the frequency spectrum at time j, with $1 \le j \le m$. Similarly each row can be seen as the evolution of frequency channel k over time, with $1 \le k \le n$.

It is assumed that the overall spectrogram $\boldsymbol{Y}$ results from the superposition of l unknown independent spectrograms Y_j. As the superposition of spectrograms is a linear operation in the time-frequency plane this yields:

$$\mathbf{Y} = \sum_{j=1}^{l} Y_j \tag{2}$$

It is then assumed that each of the Y_j can be uniquely represented by the outer product of an invariant frequency basis function f_j, and a corresponding invariant amplitude envelope or weighting function t_j which describes the variations in amplitude of the frequency basis function over time. This yields

$$\mathbf{Y}_j = f_j t_j^T \tag{3}$$

Summing the $\mathbf{Y}_j$ yields

$$\mathbf{Y} = \sum_{j=1}^{l} f_j t_j^T \tag{4}$$

In practice the assumption that the frequency basis functions are stationary means that no change in pitch can occur within the spectrogram. Casey and Westner overcome this assumption by breaking the signal into smaller blocks, inside of which the pitch can be considered stationary. However when dealing with sources that can be assumed to be stationary in pitch, such as most drum sounds, this step can be removed.

The independent basis functions correspond to features of the independent sources, and each source is composed of a number of these independent basis functions. The basis functions that compose a sound source form a low-dimensional subspace that represents the source. The basis functions that compose a source are then grouped together using a mean-field clustering algorithm. Once the low-dimensional subspaces have been identified the independent sources can be resynthesised if required.

There remains the problem of estimating the underlying basis functions to allow decomposition of the spectrogram in the manner described above. One method of doing this is Principal Component Analysis (PCA). PCA linearly transforms a set of correlated variables into a number of uncorrelated variables that are termed principal components. The first principal component contains the largest amount of the total variance as possible, and each successive principal component contains as much of the total remaining variance as possible. As a result of this property one of the uses of PCA is as a method of dimensional reduction, by discarding components that contribute minimal variance to the overall data.

One method of carrying out PCA is singular value decomposition (SVD), which decomposes $\mathbf{Y}$, an $n \times m$ matrix into

$$\mathbf{Y} = USV^T \tag{5}$$

where U is an $n \times n$ orthogonal matrix, V is an $n \times m$ orthogonal matrix and S is an $n \times m$ diagonal matrix of singular values. The columns of U contain the principal components of **Y** based on frequency, while the columns of V contain the principal components of **Y** based on time. As the number of sources p is very much smaller than n or m, we keep only the first few principal components and take these to contain our independent basis functions that describe the sources.

However PCA does not return a set of statistically independent basis functions. To obtain independent basis functions a further procedure, known as Independent Component Analysis (ICA), must be carried out [2].

Independent Component Analysis attempts to separate a set of observed signals that are composed of mixtures of a number of independent non-gaussian sources into a set of signals that contain the independent sources. The independent sources are assumed to have been mixed linearly. Using vector-matrix notation this can be stated as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{6}$$

where **x** contains the observed mixture signals, **s** contains the independent non-gaussian sources, and **A** is the mixing matrix.

To recover the independent sources ICA makes use of a corollary of the central limit theorem. The central limit theorem states that mixtures of non-gaussian signals will tend towards a gaussian distribution as the number of signals increases. As a result the mixture signals in **x** will have probability density functions that are closer to gaussian than the source signals in **s**. From this it can be seen that the original sources will have probability density functions that are more non-gaussian than any mixture of the sources. Therefore finding an unmixing matrix that gives a set of signals that are as non-gaussian as possible given the data in the mixtures will in most cases result in the recovery of the independent sources.

It should be noted that ICA cannot recover the signals at their original amplitudes or in the order in which the signals are presented. However in practice these restrictions do not affect the usefulness of ICA methods. There are numerous algorithms publicly available for performing ICA, such as FastICA and Jade [3,4]. Good reviews of ICA methods can to be found in [2,5].

ICA is performed on the basis functions that have been retained from the PCA step to yield a set of independent basis functions. It should be noted that the basis functions retained can be taken from either U or V. If taken from U the basis functions obtained after ICA will be independent in frequency. Similarly if taken from V the basis functions obtained will be independent in time. Once the independent basis functions have been obtained the corresponding amplitude envelopes or frequency basis functions can be obtained from matrix multiplication of the pseudo-inverse of the independent basis functions with the original overall spectrogram. Once these have been obtained a spectrogram of an independent subspace can be obtained as shown in equation (3). As ISA works on the magnitudes of the STFT coefficients there is no phase information available to allow resynthesis. A fast but crude way of obtaining phase information is to reuse the phase information from the original STFT. However the quality of the resynthesis using this method varies widely from signal to signal.

1.2. Optimal Information for Source Separation

Estimating the optimal amount of information to keep remains a problem. The amount of information contained in a given number of basis functions can be estimated from the normalised cumulative sum of the singular values. A threshold can then be set for the amount of information to be retained, and the following inequality can be used to solve for the number of basis functions required:

$$\frac{1}{\sum_{i=1}^{n}\sigma_i}\sum_{i=1}^{\rho}\sigma_i \geq \phi \tag{7}$$

where σ_i is the singular value of the i^{th} basis function, ϕ is the threshold and ρ is the required number of basis functions.

There is a trade-off between the amount of information to retain and the recognisability of the resulting features. Setting ϕ =1 results in a set of basis functions which support a small region in the frequency range. When ϕ << 1, the basis functions are recognisable spectral features with support across the entire frequency range. It is this case which is of interest in determining independent subspaces which represent features of the source signals.

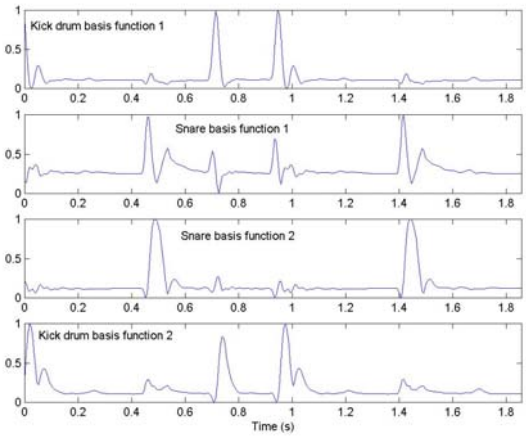

Figure 1. *ISA of drum loop (4 basis functions)*

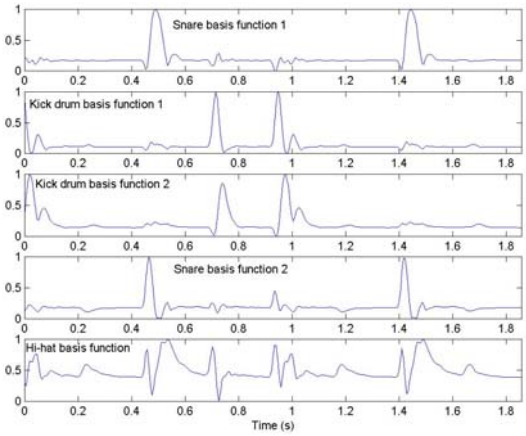

Figure 2. *ISA of drum loop (5 basis functions)*

1.3. Limitations of Independent Subspace Analysis

While ISA does provide an effective means of separating sound mixtures it should be noted that there are a number of problems with ISA. These are discussed below from the point of view of separating and transcribing drums.

The first problem is that the amount of information that needs to be retained following the PCA step for successful separation varies depending on the frequency characteristics of the sounds and their relative amplitudes. In testing the ISA method using input signals containing mixtures of three drums the number of basis functions required to effectively separate the drums was found to vary from 3-6 basis functions. Using the threshold method described previously did not always result in the correct separation of the test signals. Too low a threshold resulted in missing sources, too high a threshold resulted in the recovery of spectral features which were not usable for the purposes of drum transcription.

The problem of estimating the required information is illustrated in Figures 1 & 2. The figures show the amplitude envelopes obtained from performing ISA on a drum loop containing snare, kick drum and hi-hats. Figure 1 shows the result obtained from keeping 4 basis functions, and Figure 2 shows the result obtained from keeping 5 basis functions. As can be seen above, retaining an extra basis function allows the separation of the hi-hats. The indeterminacy in the number of basis functions required for a given separation affects the robustness of any drum transcription system using ISA, and means that the presence of an observer is required to identify the correct number of basis functions required for separation of the drums.

Secondly, as drums are broadband noise based instruments there are regions of overlap between the sounds, and as a result sometimes other drums show up as small peaks in the amplitude envelopes of the separated drums. However when good separation is obtained a simple thresholding operation is usually sufficient to identify the required events.

The quality of separation also depends on the length of the signal input. For instance a signal containing just one hi-hat and snare played simultaneously will not separate correctly. For the hi-hat/snare separation 2-4 events are typically required, depending on the frequency and amplitude characteristics of the drums used.

The method also has limitations on the number of sources it can recover, working best on signals with less than five sources. This is a result of the trade-off between the need to keep more information to allow recovery of the sources, and the loss of recognisability of the features recovered as the amount of information retained increases. However in most cases the number of drums occurring in the segment analysed will be less than five.

As can be seen from the above there are a number of limitations in the ISA method. However once these limitations are taken into account ISA provides an effective means of overcoming the masking problem encountered by Sillanpää et al when trying to identify mixtures of drums [6].

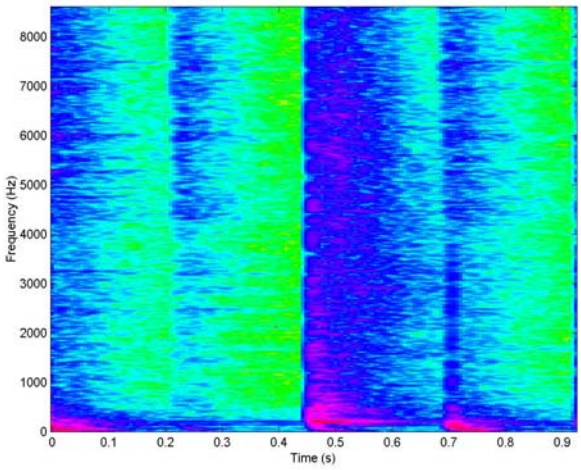

Figure 3. *STFT of a section of a drum loop*

2. SUB-BAND INDEPENDENT SUBSPACE ANALYSIS

2.1. Motivation

As noted previously the number of basis functions required to separate the sources varies depending on the frequency characteristics and relative amplitudes of the sources present. To overcome this problem it is proposed to add a sub-band processing step to the ISA method.

The addition of sub-band processing to the ISA method is motivated by observing some general properties of drums as used in popular music. The drums in a standard rock kit can be divided into two types, drums where a skin is struck, including snares, toms, and kick drums, and drums where metal is struck, including hi-hats and cymbals. The skinned drums have most of their energy in the low end of the frequency range, below 1kHz and the metal drums have most of their energy spread out over the spectrum above 2 kHz. This is illustrated in Figure 3, where the intense regions below 1 kHz correspond to the occurrence of skinned drums. Also in most popular music the skinned drums are mixed louder in the recordings than the metal drums. This means that the skinned drums dominate in ISA analysis of the input signals.

It is proposed to make use of the frequency characteristics of the drums to improve the robustness of the ISA method for transcription purposes by using sub-band processing. The signal is split into two bands, a low pass band for transcribing the skinned drums, and a high pass band for the metal drums. The low pass filter has a cutoff frequency of 1kHz, and the high pass filter has a cutoff frequency of 2 kHz. The high pass filter has the effect of removing a large amount of the energy of the skinned drums, thus allowing the metal drums to be identified with greater ease.

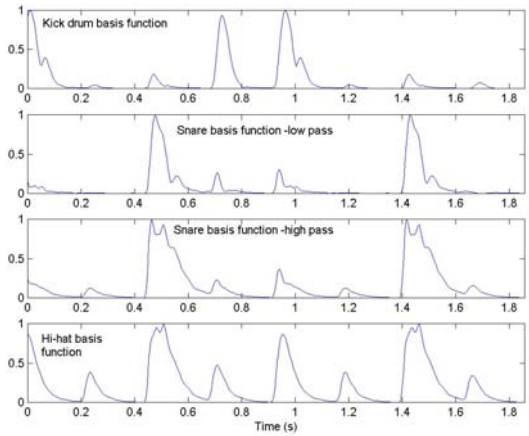

Figure 4. *Sub-band ISA of drum loop*

2.2. Drum Transcription using Sub-band ISA

To demonstrate the robustness of sub-band ISA a simple drum transcription system was implemented in Matlab. The system is limited, but effective within the confines of its limitations. It contains no explicit models of the drum types and contains no rhythmic models, but does make a number of assumptions. Firstly it is assumed that only three drums are present in the test signals, snare drums, kick drums and hi-hats. The basis for this assumption is that the basic drum patterns found in popular music consist largely of these three drums. Secondly it is assumed that the hi-hat occurs more frequently than the snare drum. Again this assumption holds for most drum patterns in popular music. Thirdly it is assumed that the kick drum has a lower spectral centroid than the snare drum. This assumption is justified in that snare drums are perceptually brighter than kick drums, and the brightness of sounds has been found to correlate well with the spectral centroid [7]. The use of sub-band processing ensures that only two basis functions are required in each band to separate the components.

Analysis starts with the signal being filtered into two bands as described previously. The low-pass signal is then passed to the ISA algorithm with only two basis functions kept from the PCA step. The spectral centroids of the separated components are calculated, and the component with the lowest centroid identified as the kick drum. The other component is then identified as the snare. As separation of the sounds is not perfect the amplitude envelopes are normalised and all peaks above a threshold are taken as an occurrence of a given drum. Onset times were calculated using a variation of the onset detection algorithm proposed by Klapuri [8]. The high-pass signal is processed in a similar manner, with the hi-hat determined as the basis function that has the most peaks in amplitude over the threshold. The remaining basis function contains the high frequency energy from the snare drum that has not been removed in filtering. Figure 4 shows the performance of sub-band ISA on the same drum loop used in figures 2 & 3. As can be seen sub-band ISA gives the required separation using only 4 basis functions, and results in much clearer separation of the hi-hats than ISA using 5 basis functions.

3. RESULTS

The system was tested on 15 drum loops containing snares, hi-hats and kick drums. The drums were taken from various sample CDs and were chosen to cover the wide variations in sound within each type of drum. The drum patterns used are examples of commonly found patterns in rock music, as well as variations on these patterns. The tempos used ranged from 80bpm to 150 bpm and different meters were used, including 4/4, 3/4 and 12/8. Relative amplitudes between the drums were varied between 0 dBs to –24 dBs to cover a wide range of situations and to make the tests as realistic as possible. The same set of analysis parameters was used on all the test signals. The results of the tests are summarized in Table 1.

Type	Total	Undetected	Incorrect	%
Snare	21	0	2	90.5
Kick	33	0	0	100
Hats	79	6	6	84.8
Overall	133	6	8	89.5

Table 1. *Drum Transcription Results.*

All the kick drums and snare drums were correctly identified, but two of the kicks were also categorized as snares. The undetetected hi-hats were in fact separated correctly but were just below the threshold for identification. Six snare hits were also identified as hi-hats due to imperfect separation. It is observed that there is a trade-off in setting the threshold level between detecting low amplitude occurrences of a drum and between incorrectly detecting drums due to imperfect separation. The threshold used was found to represent a good balance between the two. It should be noted that this level of success was achieved without the use of rhythmic models of basic drum patterns.

Due to the limitations in the time resolution of the STFT, and also due to smearing in time from overlapping windows, the detection of onset times had an average error of 10ms. It should be noted that this error tended to be consistent across all the drums in a given loop, so that inter-onset intervals remained consistent within a given loop. However it is still desirable to improve the accuracy of onset detection in sub-band ISA.

4. CONCLUSIONS AND FUTURE WORK

This paper has introduced the concept of sub-band ISA as a means of resolving the optimal information of ISA for the purposes of drum transcription. The effectiveness of this approach was demonstrated using a limited drum transcription system.

It is proposed to extend this work by incorporating drum models to generalise the drum transcription system and remove the limitations currently imposed. It is also proposed to extend the system to allow drum transcription in the presence of pitched

instruments, and to improve the accuracy of the onset detection in sub-band ISA.

5. REFERENCES

[1] Casey, M.A. & Westner, A., "Separation of Mixed Audio Sources By Independent Subspace Analysis" in Proc. Of ICMC 2000, pp. 154-161, Berlin, Germany.

[2] A. Hyvärinen and E. Oja. "Independent Component Analysis: Algorithms and Applications". *Neural Networks*, 13(4-5): pp 411-430, 2000.

[3] FastICA package for Matlab, http://www.cis.hut.fi/projects/ica/fastica/index.shtml

[4] Jade algorithm for ICA, http://www.tsi.enst.fr/icacentral/algos.html

[5] Cardoso, J.F., Blind Signal Separation: statistical Principles, Proceedings of the IEEE, Vol.9, No. 10, pp. 2009-2025, Oct 1998,1

[6] Sillanpää, Klapuri, Seppänen, Virtanen. "Recognition of acoustic noise mixtures by combining bottom-up and top-down processing". In proc. European Signal Processing Conference, EUSIPCO 2000

[7] Gordon, J., and Grey, J. M., "Perceptual Effects of Spectral Modifications on Orchestral Instrument Tones." *Computer Music Journal*, Vol. 2, N° 1, pp. 24-31, 1978

[8] Klapuri. "Sound Onset Detection by Applying Psychoacoustic Knowledge". IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 1999.